# Unsupervised Building Extraction from High Resolution Satellite Images Irrespective of Rooftop Structures

**Lizy Abraham**
*Dept. of Electronics & Communication. Engg.* *lizytvm@yahoo.com*
*LBS Institute of Technology for Women*
*Trivandrum, Kerala, India*

**Dr.M.Sasikumar**
*Head of the Department* *drmsasikumar@ yahoo.com*
*Marian Engineering College*
*Trivandrum, Kerala, India*

**Abstract**

Extraction of geospatial data from the photogrammetric sensing images becomes more and more important with the advances in the technology. Today Geographic Information Systems are used in a large variety of applications in engineering, city planning and social sciences. Geospatial data like roads, buildings and rivers are the most critical feeds of a GIS database. However, extracting buildings is one of the most complex and challenging tasks as there exist a lot of inhomogeneity due to varying hierarchy. The variety of the type of buildings and also the shapes of rooftops are very inconstant. Also in some areas, the buildings are placed irregularly or too close to each other. For these reasons, even by using high resolution IKONOS and QuickBird satellite imagery the quality percentage of building extraction is very less. This paper proposes a solution to the problem of automatic and unsupervised extraction of building features irrespective of rooftop structures in multispectral satellite images. The algorithm instead of detecting the region of interest, eliminates areas other than the region of interest which extract the rooftops completely irrespective of their shapes. Extensive tests indicate that the methodology performs well to extract buildings in complex environments.

**Keywords:** PSMS Image, Rooftop Detection, Otsu's Thresholding, Area Analysis.

## 1. INTRODUCTION

Cartographic feature extraction systems can be categorized by the type of sensor data used. Some of the researches concentrate on the fusion of more than one data sources. These methods usually use the advantages of height information in 3-d data set. One common approach is to use more than one aerial or satellite images and getting the height information using photogrammetric calculations [1, 2, 3, 4]. Also usage of new technologies such as LIDAR (Light Detection And Ranging), which provides high vertical accuracy and high point density, becomes popular. Some of these make use of the fusion of LIDAR and satellite image [5, 6]. Another important category of the Cartographic feature extraction systems extracts objects by using monocular aerial or satellite imagery. Wei and Zhao [7] introduce an approach, where they first cluster of the satellite image using an unsupervised learning method and used the shadow information to verify the existence of building. Then, for each building boundary, Canny operator is used for extracting edges and finally the system detects lines using Hough transform. Mayungaa [8] works on an active contour model which is commonly known as a snake algorithm for a semiautomatic building extraction method. In this method the user has to click to the approximate center of each building; then the algorithm generates the border of this building. Jin and Davis [9] proposed an automated building-extraction strategy for high-resolution satellite imagery that utilizes structural, contextual, and spectral information. The system runs automatically without pre-classification or any training sets, although some initial algorithm parameters must be set by the user.

Recent researches in this area focus on automatic and unsupervised extraction of buildings. Akçay and Aksoy [10] proposed a method for unsupervised segmentation and object detection in high-resolution satellite images but the system performance varies depending on different rooftop structures. Aytekin and Erener [11] proposed an algorithm for automatic and unsupervised building extraction from urban environments. Better performance is ensured by the method but the major drawback is over detection. Use of local feature vectors and a probabilistic framework for the extraction of buildings having diverse characteristics and appearance is also discussed [12, 13]. Though the method is efficient the algorithm is not strictly unsupervised. Recently, a novel approach for automatic detection of buildings with a gable roof from very-high-resolution aerial images, covering particularly rural areas is proposed [14]. The method can be modified for other rooftop structures but a single algorithm is not sufficient for detection of buildings irrespective of rooftop structures.

Most of the works in the literature are either designed for specific applications or need some prior knowledge, such as human's interaction for the extraction of buildings. Recent works which are focussed on unsupervised and automatic detection techniques are mostly restricted to specific types of shapes or surface features. A complex urban environment includes various shapes and surface materials which make the detection process complicated and in many cases pixels belonging to roof tops of buildings may wrongly identify as road pixels because both have linear features. Depending on roof top structure the system performance varies drastically and a single algorithm which is fully automatic and unsupervised which can be applied for any type of roof top structures with any complexity levels is difficult. A solution for this problem is proposed in this paper by eliminating inhomogeneities due to varying hierarchy. The algorithm, instead of focussing the region of interest, considers regions other than the areas having building features. The method is evaluated with various qualitative and quantitative measures which validates the superior performance of the proposed method.

The paper is organized as follows: The methodology and the algorithm are discussed in Section 2. Result analysis and performance evaluation are described in Section 3. Finally the paper is concluded in Section 4.

## 2. THE METHODOLOGY

In this paper, Quickbird satellite imagery is used to test the building extraction strategy. The input Pan sharpened multi-spectral satellite (PSMS) imagery is derived by pan sharpening from a low resolution(2.4m) multi-spectral (R,G,B,NIR) bands and high resolution (0.6m) panchromatic band Quick bird image. Pan sharpening is a pixel level fusion technique used to increase the spatial resolution of the multi spectral image using spatial information from the high resolution panchromatic image while preserving the spectral information in the multispectral image. Thus the input PSMS image obtained will have a resolution of 0.6m with multispectral (R, G, B, NIR) bands. This is done using Multispec32 which is a freeware multispectral image data analysis system. Multispec32 generates the .lan file of the PSMS image using high resolution panchromatic and low resolution multispectral satellite imagery. The resultant PSMS image has the same resolution as that of PAN image (Figure 1).

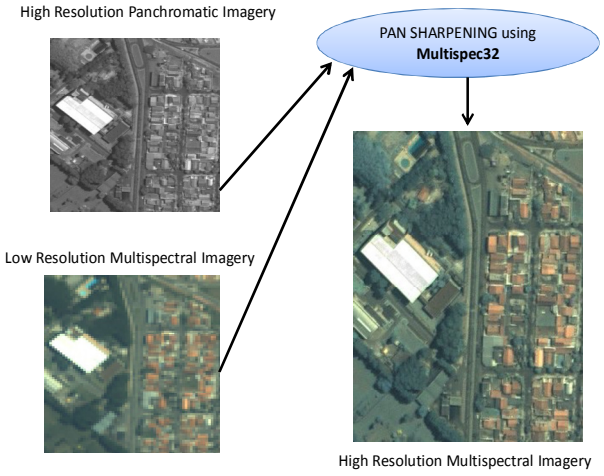


**FIGURE 1:** Pan Sharpening of Satellite Image

The method first calculates NDVI (Normalized Difference Vegetation Index) and chromaticity to intensity ratio for the initial level of segmentation. Next, rooftops and roads are detected and eliminated. Then principal component analysis and area analysis is done to get accurate results. The block diagram of the proposed system is shown in figure 2.

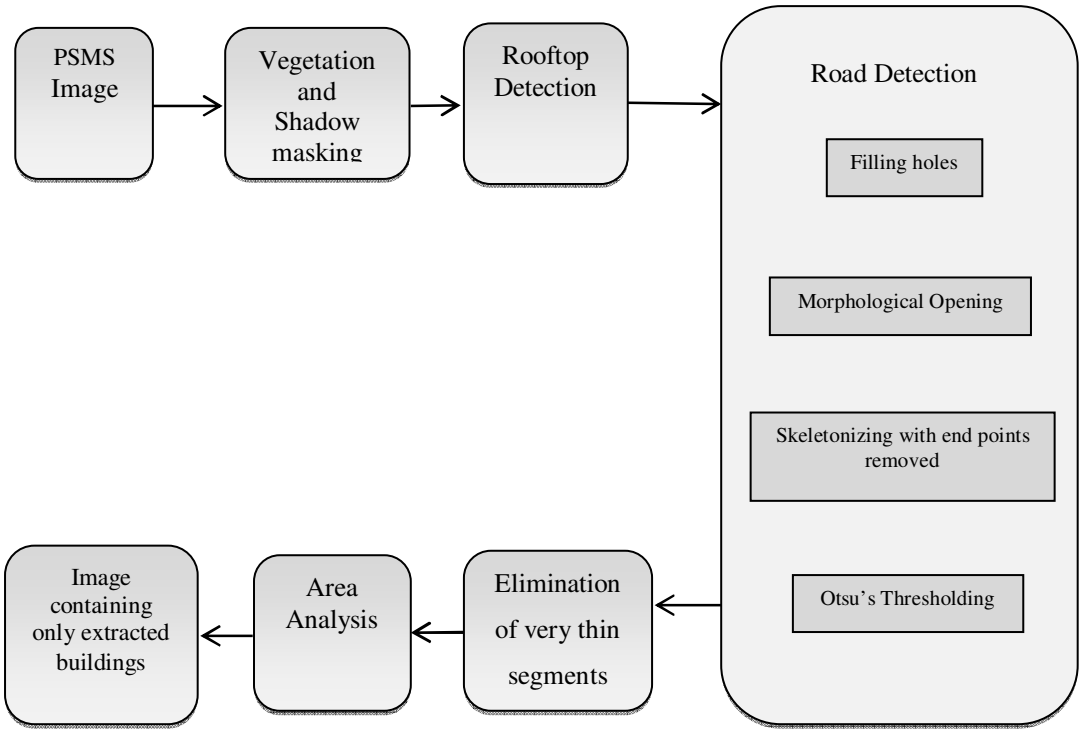


**FIGURE 2:** Overall Flow of the System

### 2.1. Calculation of Normalized Difference Vegetation Index

The NDVI is a simple graphical indicator that can be used to analyze remote sensing measurements, typically but not necessarily from a space platform, and assess whether the target being observed contains live green vegetation or not. Healthy vegetation absorbs most of the visible light that hits it, and reflects a large portion of the near-infrared light. Unhealthy or

sparse vegetation reflects more visible light and less near-infrared light. It is possible to measure the intensity of light coming off the Earth in visible and near-infrared wavelengths and quantify the photosynthetic capacity of the vegetation in a given pixel (one pixel is 0.6 square km in our concern) of land surface. Nearly all satellite Vegetation Indices employ this difference formula to quantify the density of plant growth on the Earth — near-infrared radiation minus visible radiation divided by near-infrared radiation plus visible radiation. The result of this formula is called the Normalized Difference Vegetation Index (NDVI). Written mathematically, the formula is:

$$NDVI = \frac{NIR - R}{NIR + R} \tag{1}$$

where NIR is the reflectance value in the near-infrared channel, and R is the reflectance value in the red channel. Calculations of NDVI for a given pixel always result in a number that ranges from minus one (-1) to plus one (+1); however, no green leaves gives a value close to zero. A zero means no vegetation and close to +1 (0.8 to 0.9) indicates the highest possible density of green leaves.

**IF** (NDVI$\geq$0.06) THEN ASSIGN Vegetation region;
**ELSE** ASSIGN Building region; (2)

The NDVI image can be obtained in MATLAB by reading the NIR data of the image from its .lan file generated using the software Multispec32. Thus vegetation can be eliminated by the comparison of obtained NDVI image and the original image. Regions are considered to be vegetation and the building hypotheses are rejected when the NDVI value is higher than 0.06. Figure 3 shows PSMS Quickbird image and corresponding vegetation masked image.

**FIGURE 3 (a):** Quickbird PSMS Image **(b)** Vegetation masked Image

### 2.1. Calculation of Chromaticity to Intensity Ratio

To detect shadow in the PAN sharpened, vegetation masked image, the ratio chromaticity to luminescence has to be found [15]. For this RGB space is converted to YIQ space. As the shadow regions are comparatively darker compared to other regions they will have higher ratio of Q to I. A suitable threshold is determined by Otsu's method [16] as the amount of shadows present in the high-resolution imagery will vary depending on the sun azimuth and elevation angles and the sensor azimuth angle.

Otsu's thresholding method involves iterating through all the possible threshold values and calculating a measure of spread for the pixel levels each side of the threshold, i.e. the pixels that either fall in foreground or background. The aim is to find the threshold value where the sum of

foreground and background spreads is at its minimum. This can be achieved by finding a threshold with the maximum *between class* variance and minimum *within class* variance. This can be calculated by the following equations (3 and 4).

$$\textit{Within}\ \text{Class Variance}\ \sigma_w^2 = \mathrm{W}_b\sigma_b^2 + \mathrm{W}_f\sigma_f^2 \tag{3}$$

$$\begin{aligned}\textit{Between}\ \text{Class Variance}\ \sigma_B^2 &= \sigma^2 - \sigma_w^2 \\ &= \mathrm{W}_b(\mu_b\text{-}\mu)^2 + \mathrm{W}_f(\mu_f\text{-}\mu)^2 \ (\text{where}\ \mu = \mathrm{W}_b\mu_b + \mathrm{W}_f\mu_f) \\ &= \mathrm{W}_b\mathrm{W}_f(\mu_b\text{-}\mu_f)^2 \end{aligned} \tag{4}$$

where weight, mean and variance are represented by *w,* $\mu$ and $\sigma$. The foreground and background regions are distinguished as *f* and *b*. Figure below shows the result of shadow masking.

**FIGURE 4:** Shadow masked Image

### 2.2. Rooftop Detection

For the detection of rooftops a segmentation algorithm is required. Most of the segmentation algorithms work only on gray scale images but our image is a 4-band multispectral image. In order to retain the spectral features even after segmentation a mean shift algorithm is used [17]. As for image segmentation, the aim is to cluster pixels sharing a similarity in pixel values. For this purpose, the filtering procedure is run and all convergence points are stored. The mean shift vector always points toward the direction of the maximum increase in the density. The mean shift procedure, obtained by successive computation of the mean shift vector $m_h(x_t)$ and translation of the window $x_{t+1} = x_t + m_h(x_t)$ is guaranteed to converge to a point where the gradient of density function is zero. Coming to our concern, iterative calculation of mean shift vectors converges to a stationary point of the density, which corresponds to the modes of the image, i.e. homogenous structures in general. The pixel points converging to the same mode, are closer to each other in terms of spatial extend and color bandwidth. These pixels are segmented as the same cluster. In fact, mean shift vectors are aligned towards the similarity of colors incorporating spatial information. The algorithm for mean shift segmentation is as follows (Table.1):

1. Convert the image from RGB space to LUV space.
2. Select a data point (center pixel) from the image randomly (not labeled).
3. Set the value of that point as the mean.
4. Find all the data points having the same value (a value within a specified range).
5. Cluster those points.
6. Select a point from the cluster.
7. If it is within a window or specified bandwidth(radius) from the center pixel,
   Yes-label it
   No-no change
8. Go to step 5 until all the points in the cluster are evaluated.
9. Calculate the peaks of points having the same label.
10. If distance between the peaks of two labeled segments is less than or equal to bandwidth/2,
    Yes-merge them
    No-no change
11. Compute the mean shift vector of the clustered points $m_h(x)$.

$$\mathbf{m}_h(\mathbf{x}) = \frac{\sum_{i=1}^{n} \mathbf{x}_i g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)} - \mathbf{x}$$

where h is the bandwidth parameter and g is the gradient density function.

12. Translate the window to the next point $\mathbf{x}^{t+1}$.

$$\mathbf{x}^{t+1} = \mathbf{x}^t + \mathbf{m}_h(\mathbf{x}^t)$$

13. If $\mathbf{x}^{t+1} = \mathbf{x}^t$
    Yes- go to step 14
    No- go to step 2
14. If all the points are labeled
    Yes-end
    No-go to step 1.

**TABLE 1:** Algorithm for mean shift segmentation

**FIGURE 5:** Mean shift segmented Image

### 2.3. Road Detection

After segmentation, some noises seen with the buildings in segmented regions. The edges of these regions are irregular. Hole filling followed by morphological opening operation is used to remove pseudo pixels and smooth building's edge. A hole may be defined as a background

region surrounded by a connected border of foreground pixels. Let *A* denote a set of points whose elements are 8-connected boundaries, each boundary enclosing a background region (a hole). Given a point in each $X_k$ defined by (5) as a hole, the objective is to fill all the holes with ones.

$$X_k = (X_{k-1} \oplus B) \cap A^{C} \qquad k=1,2,3,..... \tag{5}$$

Opening operation is of the form (6):

$$X \circ B = (X \Theta B) \oplus B \tag{6}$$

That is, image *X* is eroded by structure element *B* (equation 7), then it is dilated by *B*. We define,

$$B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{7}$$

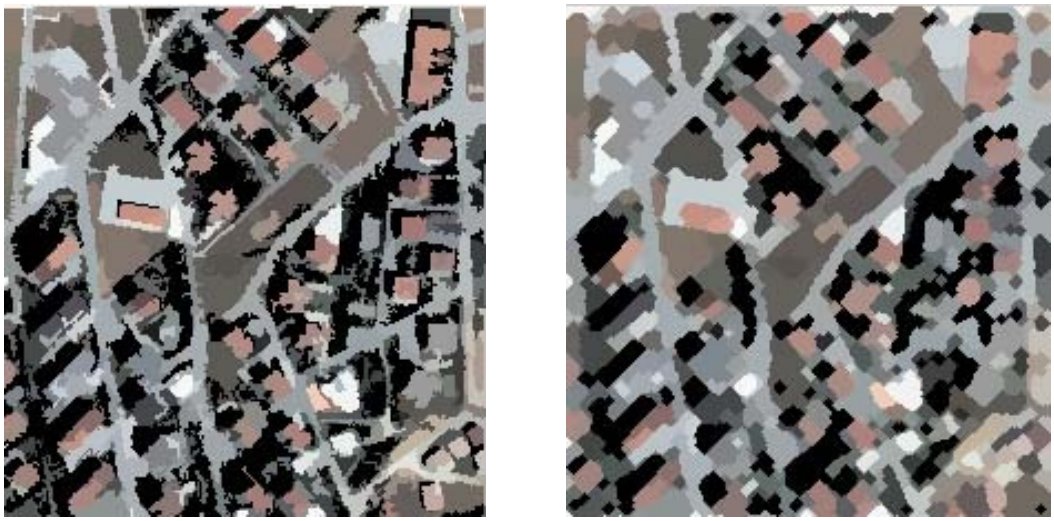

**FIGURE 6 (a):** Image after hole filling **(b)** Image after morphological opening

Figure 6(a) and 6(b) show the results of hole filling and morphological opening. After morphological treatment, edges have become smooth, but there are still some discrete noise. Pixels belonging to rooftop of buildings may wrongly identify as road pixels because both have linear features. To eliminate this, skeletonise the image (figure 7(a)), as now our region of interest is main roads alone.

For skeletonization, label pixel p if and only if the rules 1, 2, 3, 4 are all satisfied.

*Rule 1: The pixel under consideration must presently be black. If the pixel is already white, no action needs to be taken.*
*Rule 2: At least one of the pixel's close neighbours must be white.*
*Rule 3: The pixel must have more than one black neighbour.*
*Rule 4: A pixel cannot be removed if it results in its neighbours being disconnected.*

After each iteration the labeled pixels are deleted. The algorithm is performed until one pixel wide skeleton is obtained.

For identifying whether a segment is road or not, find out the length of the segments which is equal to the number of pixels in skeleton. From the distribution of length of segments threshold is automatically estimated using Otsu's method. If length is greater than this threshold then classify it as road. In the skeletonised image, there may be unwanted branches which are obviously non-road segments. So end points of the skeletons, which have only one neighbor, are removed (figure 7(b)) for better result. Figure 8 is the road masked image after these operations.

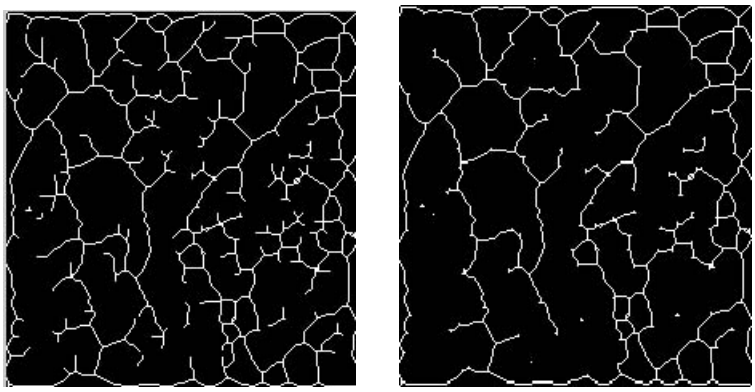

**FIGURE 7 (a):** Skeletonized Image **(b)** Skeletonized Image with end points removed

**FIGURE 8:** Road masked image

### 2.3. Elimination of Very Thin Segments

Some small objects such as cars, trees on the side of roads can also be viewed as thin stripes after skeletonization. This can be eliminated by applying, principal component analysis (PCA) to each segment. Very thin segments show large variances along the first principal component whereas small variance along the second principle component. Therefore, the ratio of the corresponding eigenvalues, provides the variances along the corresponding eigenvectors, gives an idea of how thin the segment is. Higher ratios represent unreasonably thin segments. Then, this ratio is thresholded in order to detect whether the segment is road or not. The threshold is automatically determined by Otsu's method.

Principal component analysis (PCA) is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values

of uncorrelated variables, called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has as high a variance as possible, and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (uncorrelated with) the preceding components. PCA is the simplest of the true eigenvector-based multivariate analysis.

PCA is done by calculating the mean of each co-ordinate X and Y of the segment and then finding the covariance matrix. Covariance matrix for a set of data with n dimensions is given by (8) and (9).

$$C^{m\times n} = \operatorname{cov}(Dim_i, Dim_j) \tag{8}$$

$$\operatorname{cov}(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})\,(Y_i - \overline{Y})}{(n-1)} \tag{9}$$

Then find the eigen values $\lambda_1,\lambda_2$ of $C_x$ using (10).

$$|C_x - \lambda I| = 0 \tag{10}$$

The eigen vector with the highest eigen value is the principal component. The eigen vectors are ordered by eigen value from highest to lowest. This gives the components in the order of significance. Form the feature vector as (11).

$$\textit{Feature vector} = (eig1\ eig2\ eig3\ \ldots.\ eign) \tag{11}$$

The final data is derived by the following equation (12).

$$\textit{Final data} = \textit{Row feature vector} \times \textit{Row Data Adjust} \tag{12}$$

Row Feature Vector is the matrix with the eigenvectors in the columns transposed so that eigen vectors are now in the rows, with the most significant eigenvector at the top. Row Adjust Data is the mean adjusted data transposed. The data items are in each column, with each row holding a separate dimension.

**2.4. Area Analysis**

After these operations, there are still a lot of regions which are similar to building areas. The size of these areas is very small compared to buildings and can be eliminated by connected component labelling. To verify the hypothesized connected components as building regions, area analysis is performed. For that, the minimum enclosing rectangle (MER) [18] of each and every connected component is found. Then rectangular fit is calculated as the area of the component divided by the area of its MER. If the rectangular fit is lower than a threshold, the connected component is rejected which reduces the percentage of over detection. The results of PCA and area analysis for the PSMS image are shown in figure 9.

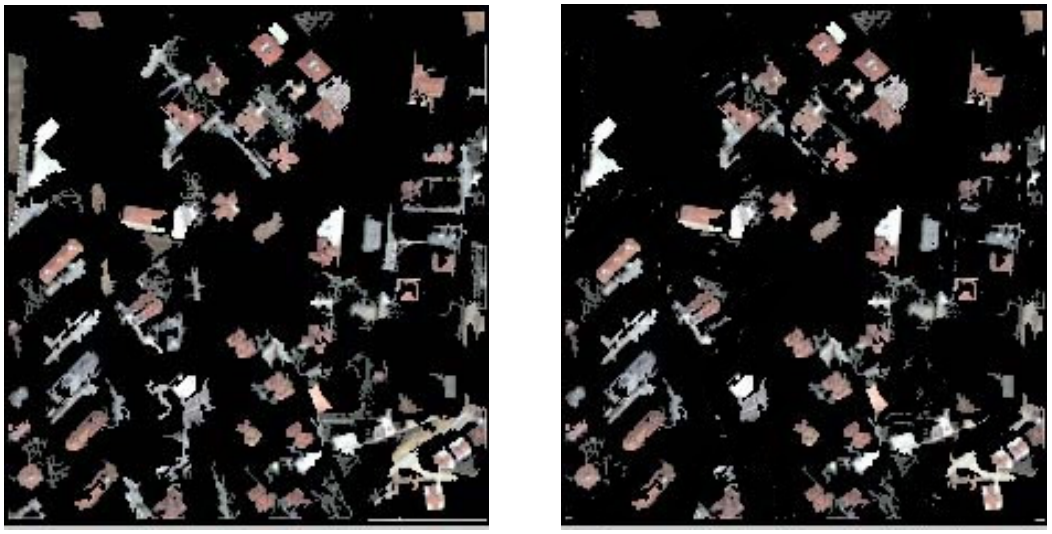

**FIGURE 9 (a) :** Image after PCA **(b)** Image after Area Analysis

## 3. RESULTS & DISCUSSION

An algorithm for fully automatic and unsupervised building extraction is presented. Figure 10 shows the extracted results compared with the manually labelled buildings used as reference. It can be seen that irrespective of shape, most of the buildings are detected without fail.

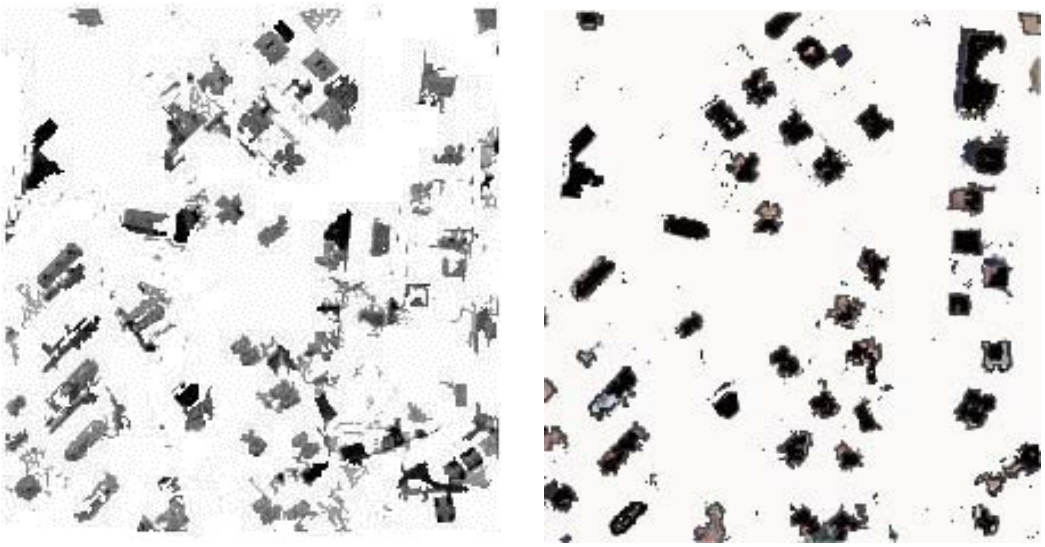

**FIGURE 10 (a):** Extracted buildings **(b)** Manually labelled buildings

Within the test area, 49 buildings were manually delineated. As seen from the output image the same number of buildings is detected but many of the scattered points can also be considered as buildings. The manually delineated buildings were used as a reference building set to assess the unsupervised building extraction accuracy. For performance evaluation, we use the evaluation measures widely accepted for building extraction [19, 20]. The extracted buildings and the manually detected buildings are compared pixel-by-pixel. The pixels in the image are categorized into four types:

*(1) True positive (TP): Both manual and unsupervised methods label the pixel belonging to the buildings.*

*(2) True negative (TN): Both manual and unsupervised methods label the pixel belonging to the background.*
*(3) False positive (FP): The unsupervised method incorrectly labels the pixel as belonging to a building.*
*(4) False negative (FN): The unsupervised method does not correctly label a pixel belonging to a building.*
Based on these categories the system performance is evaluated using the following measures:

$$\text{Branching Factor} = \frac{FP}{TP} \tag{13}$$

$$\text{Miss Factor} = \frac{FN}{TP} \tag{14}$$

$$\text{Detection Percentage} = 100 \times \frac{TP}{TP + FN} \tag{15}$$

$$\text{Quality Percentage} = 100 \times \frac{TP}{TP + FP + FN} \tag{16}$$

The *detection percentage* is the percentage of building pixels correctly labelled by the proposed method. The *quality percentage* measures the quality of the extraction process. Performance evaluations for three different images of various complexity levels are shown in Table 2. fig. 12 and fig. 13 show the building extraction results compared with the manually detected reference for image-2 and image-3 which are shown in fig. 11.

| Reference Images | Branching factor | Miss factor | Detection percentage | Quality percentage |
|---|---|---|---|---|
| Image 1 | 0.12 | 0.09 | 95.3 | 81.5 |
| Image 2 | 0.24 | 0.31 | 83.7 | 70.6 |
| Image 3 | 0.38 | 0.47 | 74.6 | 60.3 |

**TABLE 2:** Performance evaluation of extracted buildings

**FIGURE 11 (a) :** Image-2 **(b)** Image-3

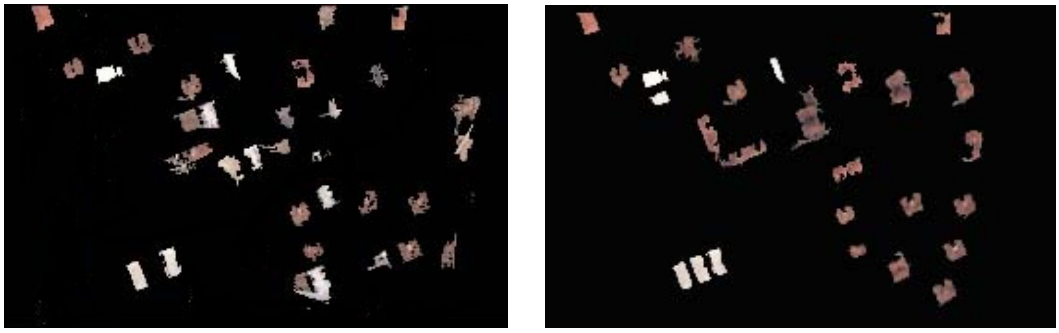

**FIGURE 12:** Extracted results compared with manually labelled reference for Image-2

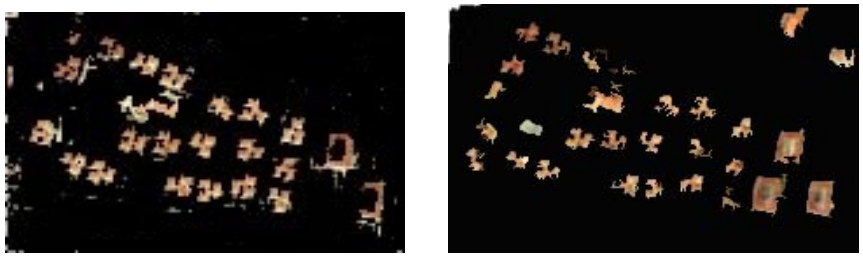

**FIGURE 13:** Extracted results compared with manually labelled reference for Image-3

The results of the performance evaluation show that the algorithm provides 81.5 % quality percentage for image 1 which is considerably good. The detection percentage for the same case is 95.3% which is very high compared to previous methods.

In our work, masking is performed on areas other than the region of interest which helped to achieve improved performance of the algorithm. It became possible to incorporate both spatial and spectral properties of the image through the use of mean shift segmentation. Principal component analysis using statistical properties helped to achieve better adaptability for different complexity levels of urban areas. Finally extensive analysis with respect to area is done, which removes most of the unwanted pixels classified as buildings. In this approach, building extraction is completely automatic and unsupervised. Here we only provide an input which is the pan-sharpened multispectral image and we get an output which is the candidate buildings.

## 5. CONCLUSION & FUTURE WORK

In this paper, a fully automatic and unsupervised system is designed for extracting high density urban area buildings irrespective of rooftop structures from satellite images. Though the variety of the type of buildings and shapes of rooftops makes the building extraction complicated, acceptable accuracy for the extracted buildings is obtained. Quantitative analysis of the work indicates that our method is better compared to existing methods. The output image includes scattered black points which may be detected as buildings. Also some roads and pavements are also detected as roads. Future work will include post processing algorithms to reduce these errors.

## 6. REFERENCES

[1] Henricsson O., “The Role of Color Attributes and Similarity Grouping in 3-D Building Reconstruction”, *Computer Vision and Image Understanding, 1998*.

[2] A. Fischer, T.H.Kolbe, F. Lang, A. B. Cremers, W. F¨orstner, L. Pl¨umer, and V. Steinhage, “Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D”, *Computer Vision and Image Understanding*, pages 185–203, 1998.

[3] T. Moons, D. Fr`ere, J. Vandekerckhove, and L. Van Gool, “Automatic modeling and 3D reconstruction of urban house roofs from high resolution aerial imagery”, In Proceedings, *Fifth European Conference on Computer Vision*, *Vol. I*, pages. 410–425, 1998.

[4] D. Koc San and M. Turker, “Automatic Building Extraction From High Resolution Stereo Satellite Images”, *ISPRS commission VII,* 2007

[5] G. Sohn and J. Dowman, "Building Extraction Using LIDAR DEMs and IKONOS Images", *ISPRS, Volume XXXIV, PART 3/W13*, 2003.

[6] F. Rottensteiner and C. Briese., “A New Method for Building Extraction in Urban Areas from High-Resolution LIDAR Data”, *Symposium der ISPRS-Comm. III. International Archives of Photogrammetry and Remote Sensing, Volume XXXIV / 3A*, pp. 295 – 301, 2002.

[7] Yanfeng Wei, Zhongming Zhao, and Jianghong Song. “Urban building extraction from high-resolution satellite panchromatic image using clustering and edge detection”, *Geoscience and Remote Sensing Symposium*, 2005.

[8] S.D Mayungaa , Dr. Y. Zhanga, and Dr. D.J. Colemana, “Semi-Automatic Building Extraction Utilizing Quickbird Imagery”, *IAPRS Vol. XXXVI, Part 3/W24*, 2005.

[9] Jin, X. and Davis, C. H., “Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information,” EURASIP Journal on Applied Signal Processing, 2196–2206 (2005).

[10] H. Akçay and S. Aksoy, “Automatic detection of geospatial objects using multiple hierarchical segmentations,” *IEEE Trans. Geosci. Remote Sens.*,vol. 46, no. 7, pp. 2097–2111, Jul. 2008.

[11] Ö. Aytekin, A. Erener, I. Ulusoy, H. S. Düzgün, “Automatic and Unsupervised Building Extraction in Complex Urban Environments from Multi-Spectral Satellite Imagery”, 4th International Conference on Recent Advances in Space Technologies, Space for the Developing World, RAST 2009, Istanbul, Turkey, June 2009.

[12] Katartzis, A. and Sahli, H., “A stochastic framework for the identification of building rooftops using a single remote sensing image,” IEEE Transactions on Geoscience and Remote Sensing 46, 259 – 271 (2007).

[13] B. Sirmacek and C. Unsalan, "A probabilistic framework to detect buildings in aerial and satellite images", IEEE Transactions on Geoscience and Remote Sensing, Vol. 49 (1), pp. 211-221, January 2011.

[14] Hazelhoff, L., De With, P.: “Localizations of buildings with a gable roof in very-highresolution aerial images” In: Proceedings of IS&T SIE Electronic Imaging, Visual Information Processing and Communication II (2011).

[15] Tsai, 2006, “A comparative study on shadow compensation of color aerial images in invariant color models”, IEEE Transactions On Geoscience And Remote Sensing, vol. 44, no. 6, June 2006.

[16] N. Otsu, "A threshold selection method from graylevel histograms," *IEEE Trans. Sys. Man Cyber.*, vol. 9, no. 1, pp. 62-66, 1979.

[17] D. Comaniciu and P. Meer. Mean shift: "A robust approach toward feature space analysis”, IEEE Trans. Pattern Anal. Machine Intell., 24:603–619, 2002.

[18] L. da F. Costa and R. M. Cesar Jr., "*Shape Analysis and Classification: Theory and Practice"*, CRC Press, Boca Raton, Fla, USA, 2001.

[19] J. A. Shufelt, "Performance evaluation and analysis of monocular building extraction from aerial imagery," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, no. 4, pp. 311–326, 1999.

[20] D. S. Lee, J. Shan, and J. S. Bethel, "Class-guided building extraction from Ikonos imagery," *Photogrammetric Engineering and Remote Sensing*, vol. 69, no. 2, pp. 143–150, 2003.